Support Vector Machine

INSTRUCTOR: HONGJIE CHEN

JUNE 13TH 2022

Best Linear Classifier (Greatest Margin)

- Multiple possible classifier candidates

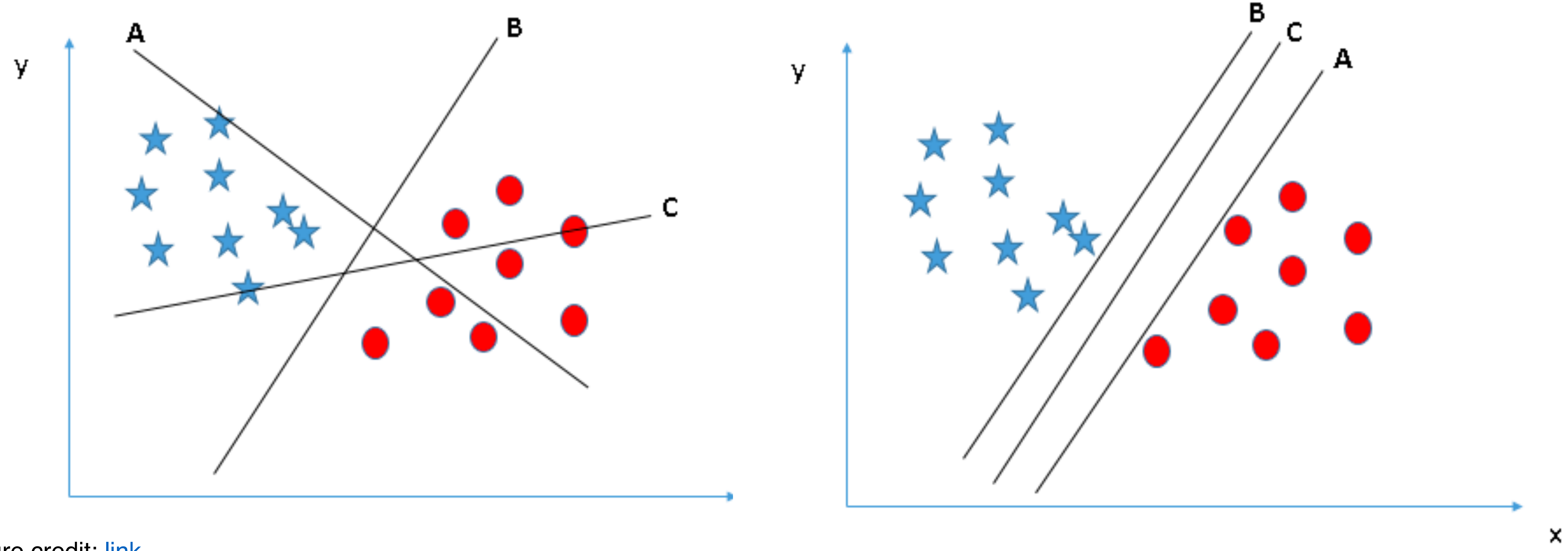
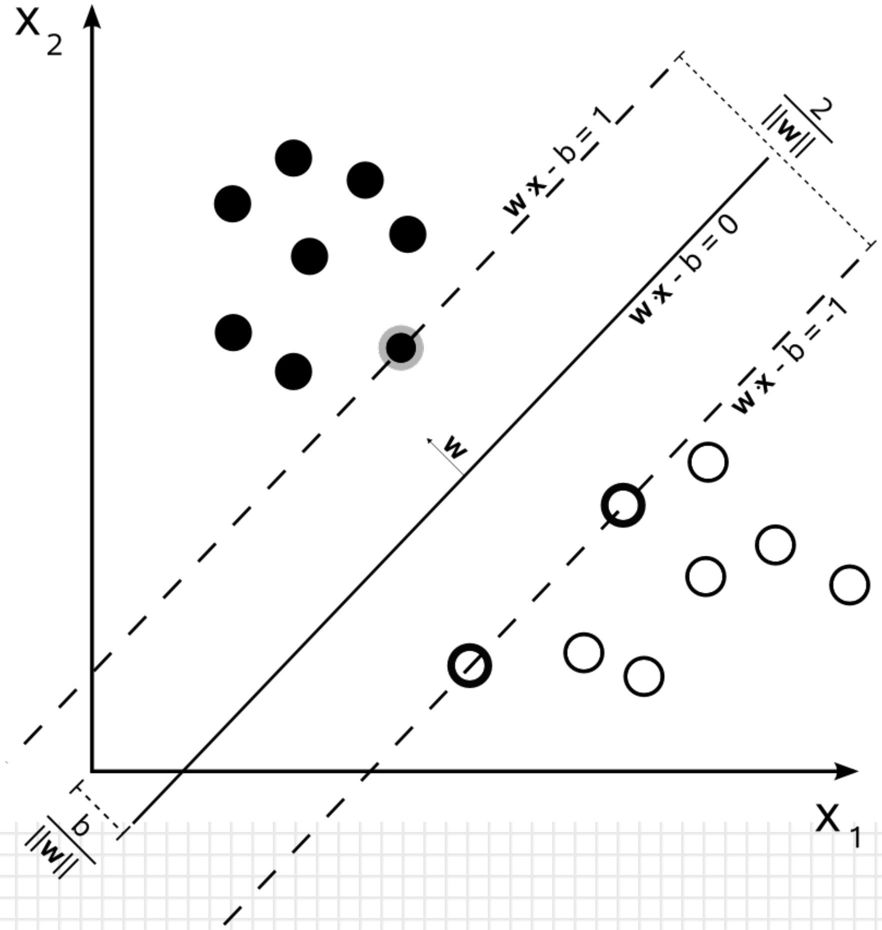


Figure credit: [link]

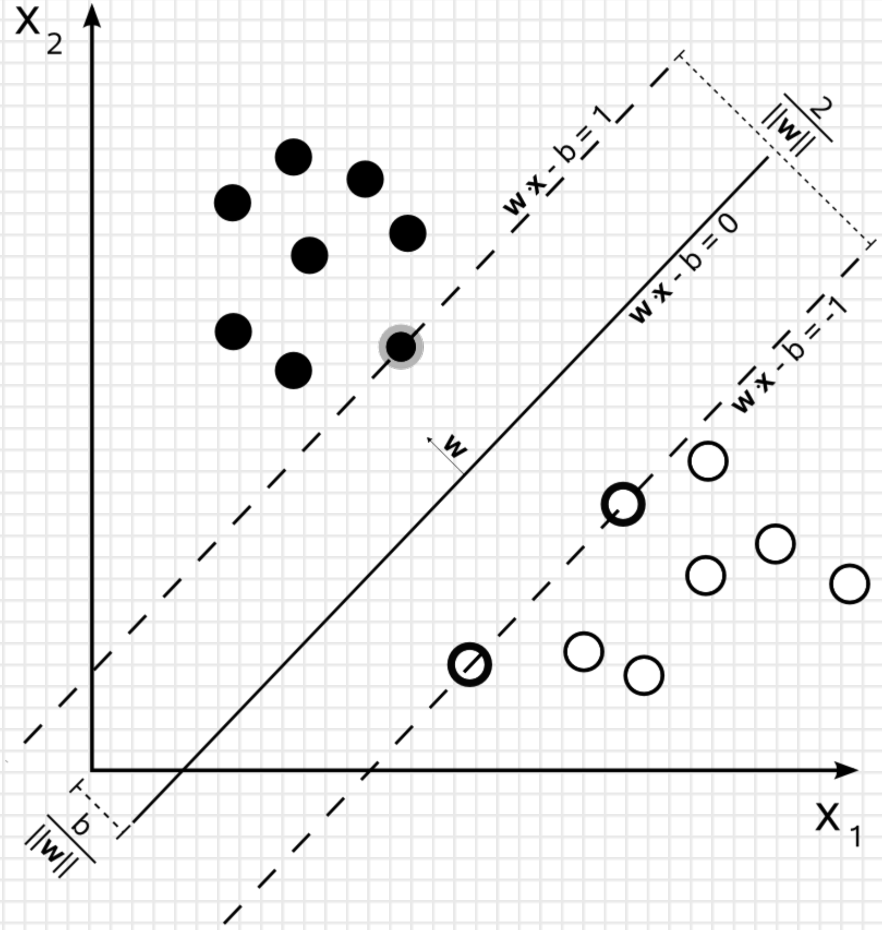
Decision Boundary

- Parameterize the decision boundary



Maximize the Margin

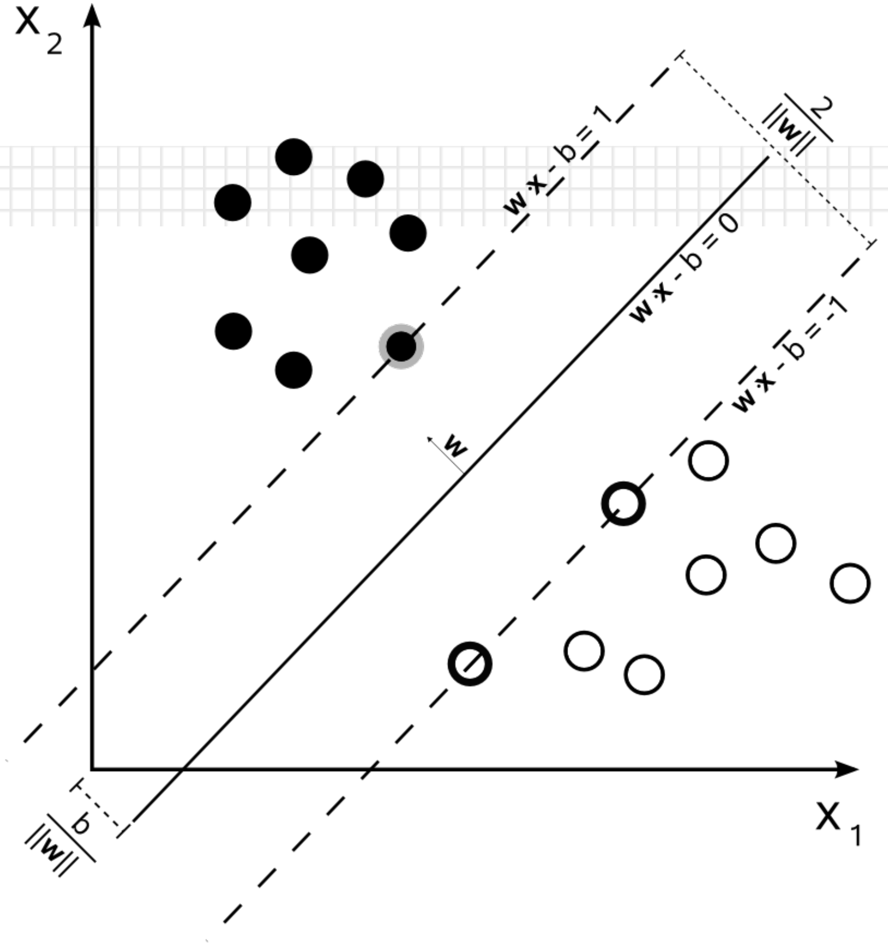
- How to obtain **w** that maximizes the margin?
- Definition of Margin?



Margin

- Margin = Distance of the closest sample(s) from the decision line/hyperplane
- Assume the margin is γ, how to represent it?

Represent γ



- Assume $\mathbf{x}_1$ on the decision line
- The closest positive sample is $(\mathbf{x}_1 + \mathbf{x}_\gamma)$
- (In the figure, assume $a = 1$)

$\mathbf{w}^T(\mathbf{x}_1 + \mathbf{x}_\gamma) + b = a$

Then $(\mathbf{w}^T\mathbf{x}_1 + b) + \mathbf{w}^T\mathbf{x}_\gamma = a$

Then $\mathbf{w}^T\mathbf{x}_\gamma = \mathbf{w} \cdot \mathbf{x}_\gamma = a$

With $\mathbf{u} \cdot \mathbf{v} = \|u\|\|v\| \cos\theta$

$\|\mathbf{w}\|\|\mathbf{x}_\gamma\| = a$

$\|\mathbf{w}\|\gamma = a \Rightarrow \gamma = \dfrac{a}{\|\mathbf{w}\|}$ Thus, $(\mathbf{w}, b) = \text{argmax}_{\mathbf{w},b}\dfrac{a}{\|\mathbf{w}\|} s.t. (\mathbf{w}^T\mathbf{x}^l + b)y^l \geq a \forall l$

VIRGINIA TECH

Support Vector Machine

- $(\mathbf{w}, b) = \text{argmax}_{\mathbf{w},b} \dfrac{a}{||\mathbf{w}||} s.t.(\mathbf{W}^T\mathbf{X}^l + b)y^l \geq a \forall l$
- a is a constant
- $(\mathbf{w}, b) = \text{argmin}_{\mathbf{w},b} \mathbf{w}^T\mathbf{w} \ s.t.(\mathbf{W}^T\mathbf{X}^l + b)y^l \geq a \forall l$
- Efficiently solved by Quadratic Programming (QP)

Primal and Dual Forms of SVM

- Primal form: solve for $\mathbf{w}, b$
- $(\mathbf{w}, b) = \text{argmin}_{\mathbf{w},b} \mathbf{w}^T \mathbf{w}$
- $s.t. (\mathbf{w}^T \mathbf{x}^l + b) y^l \geq a \forall l$
- Predict $\mathbf{x}^{new} : (\mathbf{w}^T \mathbf{x}^{new} + b) > 0$

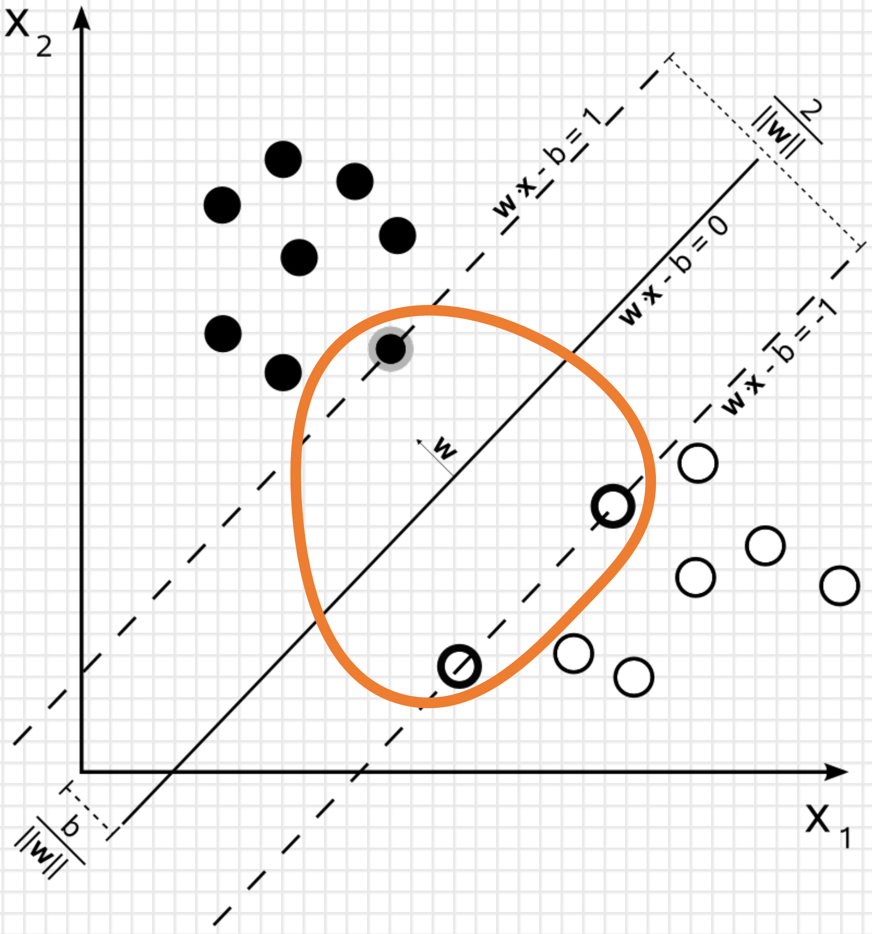
- Dual form: solve for $\alpha_1, \ldots, \alpha_L$
- $\text{argmax}_\alpha \sum_{l=1}^{L} \alpha_l - \frac{1}{2} \sum_{j=1}^{L} \sum_{k=1}^{L} \alpha_j \alpha_k y^j y^k (\mathbf{x}^{\mathbf{j}} \cdot \mathbf{x}^{\mathbf{k}})$
- $s.t. \alpha_l > 0 \forall l, \sum_{l=1}^{L} \alpha_l y^l = 0$
- Predict $\mathbf{x}^{new} : \sum_{l=1}^{L} \alpha_l y^l (\mathbf{x}^{\mathbf{new}} \cdot \mathbf{x}^{\mathbf{k}}) + b > 0$

$K(\mathbf{x}^j, \mathbf{x}^k)$

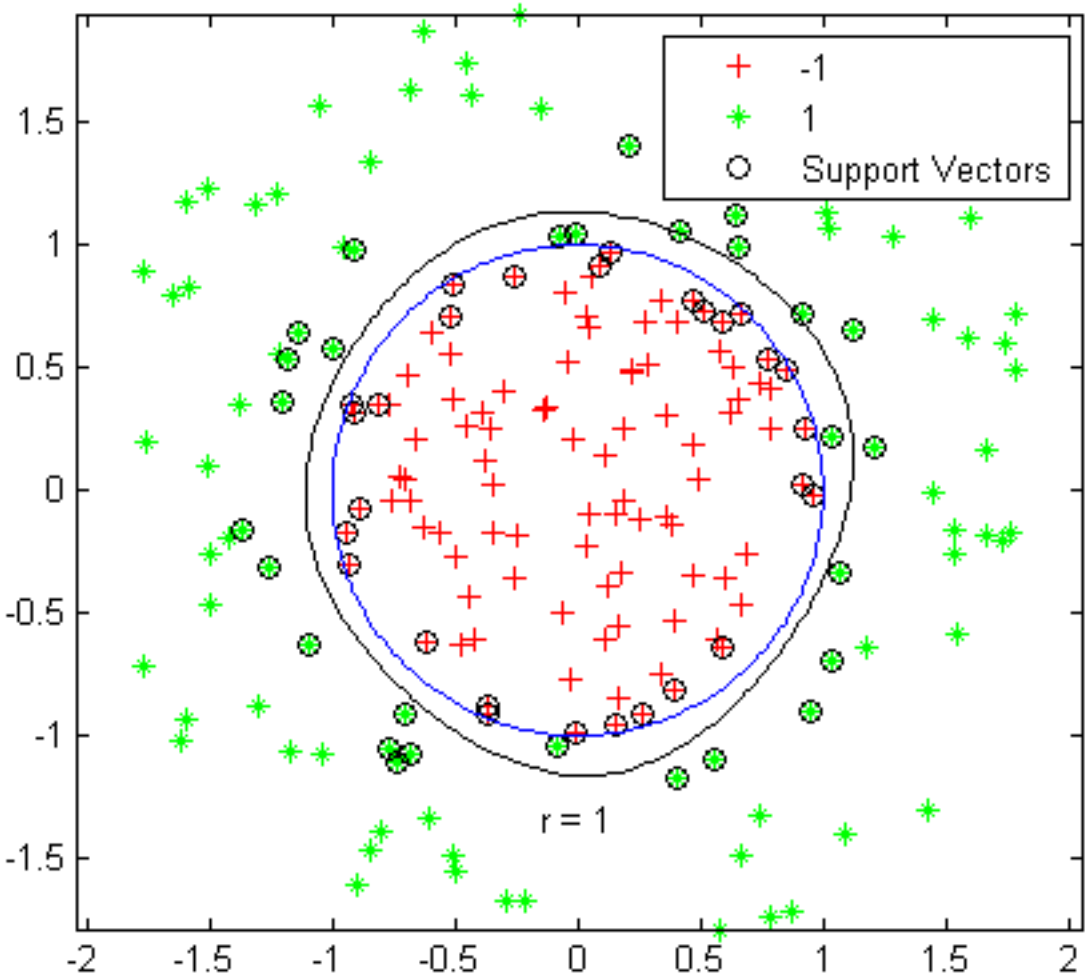
Proof*: link

Support Vectors

- The linear hyperplane is defined by **support vectors**
- Other points do not effect the decision boundary
- Only need to store the support vectors to classify new points
- What about more than two classes?



SVM with Gaussian Kernel



SVM Summary

- **Goal**: maximize margin
- Primal and dual forms
- Kernel SVM in dual form
- SVM algorithm with QP